https://www.ChinaStandards.net -> Buy the full version pdf -> Delivered by sales@ChinaStandards.net

ICS 11.040.99
CCS C 30

YY

Professional Standard of the People's Republic of China

YY/T 1833.2-2022

Artificial intelligence medical device - Quality requirements and evaluation - Part 2: General requirements for datasets

人工智能医疗器械　质量要求和评价

第 2 部分：数据集通用要求

(English translation)

Issue date: 2022-07-01　　Implementation date: 2023-07-01

Issued by the National Medical Products Administration of the People's Republic of China

https://www.chinastandards.net/standard/YYT1833.2-2022.html

https://www.ChinaStandards.net -> Buy the full version pdf -> Delivered by sales@ChinaStandards.net

Contents

Foreword I
Introduction II
1 Scope 1
2 Normative references 1
3 Terms and definitions 1
4 Dataset description requirements 4
5 Dataset quality requirements 10
6 Dataset quality conformity evaluation 16
Annex A (Normative) Dataset type description 21
Annex B (Informative) Data screening and cleaning description 23
Bibliography 25

https://www.chinastandards.net/standard/YYT1833.2-2022.html

https://www.ChinaStandards.net -> Buy the full version pdf -> Delivered by sales@ChinaStandards.net

Foreword

ChinaStandards.net is in charge of this English translation. In case of any doubt about the English translation, the Chinese original shall be considered authoritative.

This document is developed in accordance with the rules given in GB/T 1.1-2020 *Directives for standardization - Part 1: Rules for the structure and drafting of standardizing documents*.

This document is Part 2 of YY/T 1833 under the general title of *Artificial intelligence medical device - Quality requirements and evaluation*. The following parts of YY/T 1833 have been issued:

——*Part 1: Terminology*;

——*Part 2: General requirements for datasets*.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. The issuing body of this document shall not be held responsible for identifying any or all such patent rights.

This document was proposed by the National Medical Products Administration of the People's Republic of China.

This document is under the jurisdiction of the Technical Organization on AI Medical Device of Standardization Administration of China.

https://www.chinastandards.net/standard/YYT1833.2-2022.html

https://www.ChinaStandards.net -> Buy the full version pdf -> Delivered by sales@ChinaStandards.net

Introduction

In recent years, artificial intelligence medical devices have continued to develop and become an emerging trend in the field of medical device standardization. China has initially established a standard system for artificial intelligence medical devices. In the standard system, YY/T 1833 *Artificial intelligence medical device - Quality requirements and evaluation* is the basic general standard to provide guidance for standardization activities in the subdivision field, which is to be composed of eight parts.

——*Part 1: Terminology*, designed to provide terminology for the quality evaluation activities of artificial intelligence medical devices.

——*Part 2: General requirements for datasets*, designed to propose the general quality requirements and evaluation methods for datasets.

——*Part 3: General requirement for data annotation*, designed to propose the quality requirements and evaluation methods for data annotation.

——*Part 4: General requirements for traceability*, designed to propose the general requirements and evaluation methods for traceability of artificial intelligence medical devices.

——*Part 5: Security requirements for algorithm*, designed to standardize the security requirements and evaluation methods for artificial intelligence algorithms used for artificial intelligence medical devices.

——*Part 6: Environment requirements*, designed to standardize the requirements and evaluation methods for operating environmental condition of artificial intelligence medical devices.

——*Part 7: Privacy protection requirements*, designed to enhance the ability of artificial intelligence medical devices to protect receivers' privacy.

——*Part 8: Ethical requirements*, designed to realize the requirements of artificial intelligence ethics from the technical level and protect people's rights and interests.

This document provides other parts with working ideas related to dataset quality evaluation.

https://www.chinastandards.net/standard/YYT1833.2-2022.html

https://www.ChinaStandards.net -> Buy the full version pdf -> Delivered by sales@ChinaStandards.net

Datasets are commonly used resources in the development, training, testing, quality control and other aspects of artificial intelligence medical devices, which have an important impact on product quality. As Part 2 of YY/T 1833 *Artificial intelligence medical device - Quality requirements and evaluation*, this document incorporates datasets into the product quality evaluation system, and also provides a basis for the subsequent formulation of special requirements for datasets in the subdivision field.

https://www.chinastandards.net/standard/YYT1833.2-2022.html

https://www.ChinaStandards.net -> Buy the full version pdf -> Delivered by sales@ChinaStandards.net

Artificial intelligence medical device - Quality requirements and evaluation - Part 2: General requirements for datasets

1 Scope

This document specifies the general quality requirements and evaluation methods for datasets used in the full life cycle of artificial intelligence medical devices.

It is applicable to the development and evaluation of datasets used in the research and development, production, testing, quality control and other aspects of artificial intelligence medical devices.

2 Normative references

The following documents contain provisions which, through reference in this text, constitute provisions of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

GB/T 2828.4 *Sampling procedures for inspection by attributes - Part 4: Procedures for assessment of declared quality levels*

GB/T 2828.11 *Sampling procedures for inspection by attributes - Part 11: Procedures for assessment of declared quality levels for small population*

GB/T 6378.4 *Sampling procedures for inspection by variables - Part 4: Procedures for assessment of declared quality levels for mean*

YY/T 1833.1 *Artificial intelligence medical device - Quality requirements and evaluation - Part 1: Terminology*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in YY/T 1833.1 as well as the followings apply.

3.1
inspection by attributes
inspection relating to a specified requirement or a set of specified requirements that either classifies a unit product as acceptable or rejected, or only counts the number of rejected ones in unit products

[Source: GB/T 2828.1-2012, 3.1.3]

https://www.chinastandards.net/standard/YYT1833.2-2022.html

https://www.ChinaStandards.net -> Buy the full version pdf -> Delivered by sales@ChinaStandards.net

3.2
variables quality characteristic
quality characteristic of the unit product to be inspected that can be measured on a continuous scale

[Source: GB/T 8054-2008, 3.1.3]

3.3
sampling inspection by variables
process of randomly taking a certain number of unit products from the lot according to the specified sampling scheme, and then obtaining their quality characteristic values by measurement, test or other methods, comparing these values with the quality requirements, and finally judging whether the lot of products can be accepted

[Source: GB/T 8054-2008, 3.1.4]

3.4
lot
a definite part of a population composed under basically the same conditions according to sampling purposes

[Source: GB/T 10111-2008, 3.1.4]

3.5
accuracy
a measure that the data is correct in content and valid in form

[Source: GB/T 11457-2006, 2.22, modified]

3.6
precision
degree of accuracy or difference with respect to the quantities stated, such as 2-digit decimal number versus 5-digit decimal number

[Source: GB/T 11457-2006, 2.1160]

Note: The meaning of precision is different from the accuracy in algorithm performance indexes.

3.7
consistency
degree of consistency, standardization and non-contradiction among different stages or parts of the dataset

[Source: GB/T 11457-2006, 2.320, modified]

https://www.chinastandards.net/standard/YYT1833.2-2022.html

https://www.ChinaStandards.net -> Buy the full version pdf -> Delivered by sales@ChinaStandards.net

3.8
availability
degree to which a dataset is operable or available when it is put into use

[Source: GB/T 11457-2006, 2.115, modified]

3.9
information security
protection of data against accidental or deliberate access, use, modification, destruction or disclosure

security also involves the physical protection of receivers' privacy, metadata, communication and computer installation

[Source: GB/T 11457-2006, 2.1420, modified]

3.10
portability
degree to which a dataset can be installed, replaced, or moved from one system to another, while maintaining the existing quality attributes

3.11
dataset manufacture responsible organization
entity responsible for the design and manufacture of a dataset

[Source: GB 9706.1-2020, 3.101, modified]

3.12
outlier
observations in a set of data that are considered inconsistent with other data in the set

Note: Outliers may originate from different populations or may result from incorrect recording or other errors.

[Source: CNAS GL-002-2018, 3.6, modified]

3.13
dataset description
document stating various properties of a dataset

https://www.chinastandards.net/standard/YYT1833.2-2022.html